WHAT IS CLAIMED:

1. A method for determining equivalent descriptions for an information need, comprising:

identifying a list of queries issued by one or more users;

identifying a candidate pair of equivalent descriptions by locating two queries that refer to the same information need;

calculating a score for the candidate pair dependent on the frequency with which the candidate pair occurs in the list; and

determining that each half of the candidate pair is an equivalent description for the information need if the score is above a defined threshold.

2. The method of claim 1, wherein identifying a candidate pair comprises:

locating two queries that contain at least one term in common; and

identifying as a candidate pair the portions of the two queries that are not in common.

3. The method of claim 1, wherein identifying a candidate pair comprises:

identifying, in a first description, a term T1 having characters C_i, where i = 1 through n;

identifying, in a second description, a sequence of n terms, $T2_1$, $T2_2$...$T2_n$; and

determining that term T1 and terms $T2_1$, $T2_2$...$T2_n$ are a candidate pair if each C_i matches the first letter of $T2_i$.

4. The method of claim 1, wherein calculating a score comprises:

determining a first frequency with which the candidate pair occurs within the list;

determining a second frequency with which one half of the candidate pair occurs within the list; and

calculating a score based on a ratio of the first frequency and the second frequency.

5. The method of claim 1, further comprising excluding any candidate pair with a frequency of occurrence in the list below a defined threshold.

6. The method of claim 1, further comprising excluding any candidate pair wherein one half of the candidate pair contains a misspelled term.

7. The method of claim 1, further comprising excluding any candidate pair wherein it is determined that one half of the candidate pair is an alternative rather than an equivalent for the second half of the candidate pair.

8. The method of claim 7, wherein the determination comprises:

locating a collection of documents;

identifying lists within the collection, wherein each list contains both halves of the candidate pair; and

determining that one half of the candidate pair is an alternative for the second half based on the frequency with which each half occurs in the lists.

9. A method for determining equivalent descriptions for an information need, comprising:

identifying a plurality of descriptions that are associated with a plurality of information needs;

identifying a candidate pair of equivalent descriptions by locating two descriptions that refer to the same information need;

calculating a score for the candidate pair dependent on the frequency with which the candidate pair occurs in the plurality of descriptions; and

determining that each of the candidate pair is an equivalent description for the information need if the score is above a defined threshold.

10. The method of claim 9 wherein the plurality of descriptions comprises an historical log of user queries.

11. The method of claim 10, further comprising sorting the log by user.

12. The method of claim 11, further comprising sorting the log by the time when the query was issued.

13. The method of claim 9 wherein identifying a candidate pair comprises:

identifying two descriptions that contain a common term; and

identifying as a candidate pair the terms not in common between the two descriptions.

14. The method of claim 9 wherein identifying a candidate pair comprises:

comparing each letter of a term in a first description against the corresponding first letter of terms in a second description; and

determining, based on the comparison, that the term in the first description and the corresponding terms in the second description are a candidate pair.

15. The method of claim 9 wherein calculating a score comprises:

determining a first frequency with which the candidate pair occurs within the plurality of descriptions;

determining a second frequency with which one half of the candidate pair occurs within the plurality of descriptions; and

calculating a score based on a ratio of the first frequency and the second frequency.

16. The method of claim 9 wherein calculating a score comprises:

determining a first frequency with which the candidate pair occurs within the plurality of descriptions;

determining a second frequency with which one half of the candidate pair occurs within the plurality of descriptions;

determining a third frequency with which the other half of the candidate pair occurs within the plurality of descriptions;

calculating a score based on a ratio of the first frequency and the smaller of the second and third frequencies.

17. A method for determining synonyms, comprising:

obtaining a list of search queries issued by one or more users;

sorting the list first by user and second by the time when the query was issued;

selecting a set of adjacent queries for a single user;

identifying, from the set, two queries that contain at least one query term in common;

identifying as a candidate synonym pair the uncommon portions of the two queries;

calculating a score for candidate synonym pair dependent on the frequency with which the candidate synonym pair occurs in the list; and

determining that each half of the candidate synonym pair is a synonym of the other half if the score is above a defined threshold.

18. The method of claim 17, wherein calculating a score comprises:

determining a first frequency with which the candidate synonym pair occurs within the list;

determining a second frequency with which one half of the candidate pair occurs within the list; and

calculating a score based on a ratio of the first frequency and the second frequency.

19. The method of claim 17, further comprising excluding any candidate synonym pair with a frequency of occurrence below a defined threshold.

20. The method of claim 17, further comprising excluding any candidate synonym pair wherein one half of the candidate synonym pair contains a misspelled term.

21. The method of claim 17, further comprising excluding any candidate synonym pair wherein it is determined that one half of the candidate synonym pair is an alternative rather than an equivalent for the second half of the candidate synonym pair.

22. The method of claim 21, wherein the determination comprises:

locating a collection of documents;

identifying lists within the collection, wherein each list contains both halves of the candidate synonym pair; and

determining that one half of the candidate synonym pair is an alternative for the second half based on the frequency with which each half occurs in the lists.

23. A method for determining equivalent descriptions for an information need, comprising:

creating a list of anchor text units;

determining a subset of the list that refers to the same information need;

locating, within the subset, two anchor text units contain at least one term in common;

identifying as a candidate pair of equivalent descriptions the uncommon portions of the two anchor text units;

calculating a score for the candidate pair dependent on the frequency with which the candidate pair occurs in the list; and

determining that each half of the candidate pair is an equivalent description for the information need if the score is above a defined threshold.

24. An apparatus for determining equivalent descriptions for an information need, comprising:

means for identifying a list of queries issued by one or more users;

means for identifying a candidate pair of equivalent descriptions by locating two queries that refer to the same information need;

means for calculating a score for the candidate pair dependent on the frequency with which the candidate pair occurs in the list; and

means for determining that each half of the candidate pair is an equivalent description for the information need if the score is above a defined threshold.

25. An apparatus for determining equivalent descriptions for an information need, comprising:

at least one memory having program instructions, and

at least one processor configured to execute the program instructions to perform the operations of:

identifying a list of queries issued by one or more users;

identifying a candidate pair of equivalent descriptions by locating two queries that refer to the same information need;

calculating a score for the candidate pair dependent on the frequency with which the candidate pair occurs in the list; and

determining that each half of the candidate pair is an equivalent description for the information need if the score is above a defined threshold.